nature biotechnology

Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry

Xiao-Kang Lun[1,2], Vito R T Zanotelli[1,3], James D Wade[1,4], Denis Schapiro[1,3], Marco Tognetti[1,2,5], Nadine Dobberstein[1] & Bernd Bodenmiller[1]

Signaling networks are key regulators of cellular function. Although the concentrations of signaling proteins are perturbed in disease states, such as cancer, and are modulated by drug therapies, our understanding of how such changes shape the properties of signaling networks is limited. Here we couple mass-cytometry-based single-cell analysis with overexpression of tagged signaling proteins to study the dependence of signaling relationships and dynamics on protein node abundance. Focusing on the epidermal growth factor receptor (EGFR) signaling network in HEK293T cells, we analyze 20 signaling proteins during a 1-h EGF stimulation time course using a panel of 35 antibodies. Data analysis with BP-R^2, a measure that quantifies complex signaling relationships, reveals abundance-dependent network states and identifies novel signaling relationships. Further, we show that upstream signaling proteins have abundance-dependent effects on downstream signaling dynamics. Our approach elucidates the influence of node abundance on signal transduction networks and will further our understanding of signaling in health and disease.

Signaling networks are at the core of cellular information processing and transform external signals into cellular responses. Signals are transduced by modulating enzymatic activities mainly via protein phosphorylation, and cells implement sophisticated mechanisms, such as feedback loops, pathway crosstalk, and differential enzyme localization, to integrate signals and drive cellular processes and physiological outputs. The abundance of individual signaling pathway components (nodes) is central to the activity and output of a signaling network[1]. Changes in node abundance are tightly regulated and control biological programs such as stem cell differentiation and embryogenesis[2]. Abundance deregulation of particular signaling network nodes by genomic, transcriptional, or post-transcriptional regulatory defects[3–5] underlies human diseases, the prime example being cancer[6]. Copy number alterations of genes encoding critical proteins[7–9], independent of mutations that constitutively change enzymatic activity[10], drive progression of many cancer types. Genomic instability in cancer cells causes abnormally broad distributions of signaling protein abundances in a given tumor[11], yet the consequences of the protein abundance levels on signaling properties is poorly understood limiting our ability to rationally design therapies.

The EGFR signaling network is affected by gene copy number alterations that deregulate protein abundances (e.g., of EGFR, HER2, ERK, and AKT) in a number of cancer types[7–9]. EGFR signaling controls cell growth, motility, survival, differentiation, and metabolism[12]. Many drugs target the activity of the EGFR signaling network[13,14]. The receptor tyrosine kinase (RTK) function of EGFR is activated by its dimerization upon ligand binding. EGFR auto-phosphorylation recruits adaptor proteins that typically activate the MAPK/ERK and AKT signaling pathways. The MAPK/ERK branch activates the GTPase RAS, which triggers a kinase phosphorylation cascade consisting of RAF, MEK, ERK, and p90RSK. The output of the MAPK/ERK branch is transcription of genes regulating growth and division[15,16]. Signal transduction through the AKT branch starts by PI3K activation, producing PIP3, which recruits AKT and PDK1 to the plasma membrane. PDK1 phosphorylates AKT[15,17], which mediates signaling through the mTORC1 complex to modulate translation via p70S6K and 4EBP1 (ref. 17). Other AKT targets are GSK3β, PRAS40, and TSC2. The AKT pathway controls cell survival, proliferation, and migration[17]. STAT proteins and the PKC pathway can also be activated by EGFR-mediated signaling[18,19]. EGFR signaling involves crosstalk and feedback loops both internally (e.g., active ERK attenuates upstream RAF or MEK signaling via negative feedback)[15] and with other signaling pathways (e.g., WNT and TGF-β pathways)[20,21].

Classically, two approaches are used to characterize the effect of proteins on signal transduction. The first approach analyzes cell populations. Here, western blot analysis, mass spectrometry, RNA microarrays, and synthetic lethality screens are used to identify signaling relationships[22–24]. Protein–protein interaction analyses are used to determine which proteins in a network directly interact[23,25]. Population-based methods yield a comprehensive view of signaling but are difficult to use in analysis of protein abundance dependencies owing to inherent limitations. Proteins must be expressed at different abundances or cells must be sorted to yield a non-continuous abundance titration. Such methods result in a large number of samples, and cell-to-cell protein

[1]Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland. [2]Molecular Life Science PhD Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland. [3]Systems Biology PhD Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland. [4]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA. [5]Institute of Biochemistry, ETH Zürich, Zürich, Switzerland. Correspondence should be addressed to B.B. (bernd.bodenmiller@imls.uzh.ch).

Received 28 July 2016; accepted 15 December 2016; published online 16 January 2017; doi:10.1038/nbt.3770

abundance variations within each sample remain masked. The second approach studies signaling relationships in single cells. Here, fluorescence microscopy and flow cytometry are used with a variety of reporters and assays, including proximity ligation assay[26] or fluorescence resonance energy transfer[27]. These approaches allow study of signaling relationships and dynamics through time and space; however, only a few signaling nodes can be measured simultaneously.

A recently developed single-cell-analysis technology, called mass cytometry, allows for the simultaneous measurement of over 40 signaling nodes in single cells using metal-isotope-tagged antibodies[28,29]. This capability makes mass cytometry uniquely suited to comprehensively query the function of nodes in signaling networks within heterogeneous cell populations. Mass cytometry is quantitative and, in combination with mass-tag cellular barcoding, a powerful screening tool[28]. Algorithms to analyze multiplexed single-cell mass cytometry data allow quantification of signaling relationships, therefore helping to decipher the highly complex network behaviors that operate even in simple biological systems[30].

Here, we coupled protein overexpression with mass cytometry to measure the effect of varying node abundance on the activation state and signaling relationships of an unstimulated EGFR signaling network, as well as the signaling dynamics of the network in response to EGF stimulation. We exploited the finding that transient protein overexpression in a cell population typically produces a continuous abundance range of the target protein over four orders of magnitude. We overexpressed 20 central EGFR signaling network proteins individually in human embryonic kidney (HEK) 293T cells, sampled during an EGF stimulation time course over 60 min totaling 360 conditions. An average of 11,000 cells per condition was analyzed with a panel of 35 antibodies to provide a comprehensive single-cell proteomic EGFR network analysis. To identify signaling relationships in this data set, we developed a statistical measure that we call 'binned pseudo R-squared' (BP-R^2) that recapitulated known signaling relationships and identified relationships that were—to the best of our knowledge—not described previously. Thus, our experimental and computational approach enables the study of how the strength and dynamics of signal transduction are tuned by node abundances.

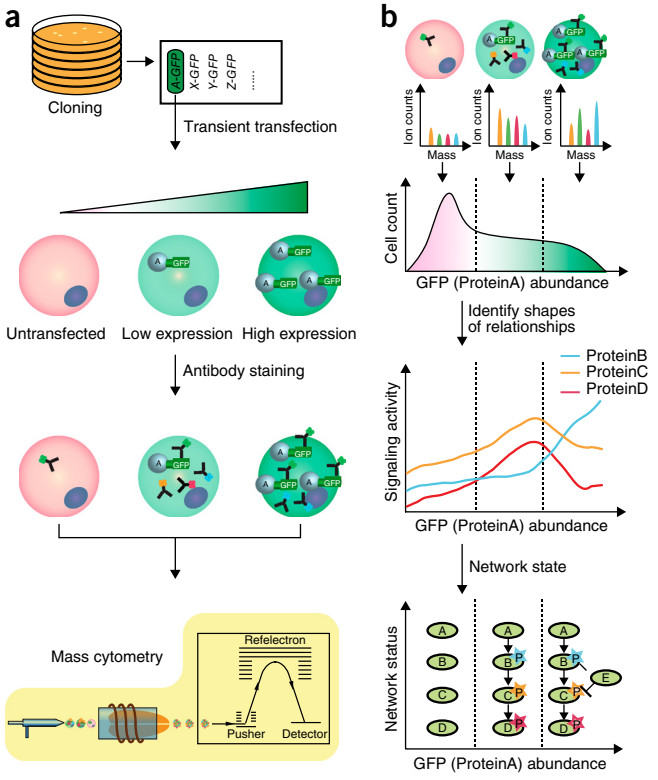

Figure 1 Workflow of abundance-dependent network analysis. (**a**) Experimental workflow. Signaling POIs are cloned into vectors containing a CMV promoter and a GFP-tag sequence to transiently overexpress GFP-tagged POIs in HEK293T cells. We quantify anti-GFP antibody as readout of POI–GFP abundance, together with other 35 markers, by mass cytometry. (**b**) Data analysis workflow. Cells were ordered based on the GFP signal, providing a continuous POI titration, which was then coupled to other signaling markers to determine the abundance dependencies of network activation state and signaling dynamics in the network after transfection. The network in the illustration does not represent an actual biological example.

RESULTS

Analyzing continuous protein abundance dependencies

To systematically identify and characterize protein-abundance-dependent signaling relationships, dynamics, and network activation states, we exploited the variation and large dynamic range of protein abundance induced by transient transfection, and used mass cytometry to quantify the abundance of the transfected protein of interest (POI) in conjunction with comprehensive signaling network readouts in single cells. We cloned genes encoding POIs into vectors containing a cytomegalovirus (CMV) promoter and a GFP-tag sequence[31] to transiently overexpress GFP-tagged POIs in HEK293T cells (**Fig. 1a**). The tagged protein abundance was measured by mass cytometry using an anti-GFP antibody (**Fig. 1a**). Ordering the measured cells based on the GFP signal provided a continuous POI titration (**Fig. 1b**). Typically, not all cells were transfected, yielding an internal control for every experiment. To measure the single-cell EGFR signaling network states, we designed and validated a panel of 35 antibodies that mostly detect phosphorylation sites on signaling proteins (**Supplementary Tables 1–3**). These data were used to determine the abundance dependencies of network activation state and signaling dynamics (**Fig. 1b**).

To validate our system we confirmed that, first, the GFP tag was reliably detected by mass cytometry (**Supplementary Fig. 1**); second, the GFP tag did not affect the localization and activity of the POI (**Supplementary Figs. 2** and **3**, **Supplementary Table 4**, and **Supplementary Dataset 1**); third, POI expression levels were linearly related to GFP abundance, validating GFP as readout of the total POI abundance (**Supplementary Fig. 4a**,**c**); fourth, POI overexpression for 18 h (i.e., the time point of our experiments) did not alter the underlying network structure (**Supplementary Fig. 4b**,**c**); fifth, the antibody-based GFP quantification by mass cytometry was comparable to fluorescence-activated cell sorting (FACS; **Supplementary Fig. 5**); sixth, the cell culture media and cell detachment did not alter signaling processing in the EGFR network (**Supplementary Figs. 6** and **7**); and, seventh, the levels of the GFP-tagged POIs were stable during the 1-h EGF stimulation time course (**Supplementary Fig. 8** and **Supplementary Video 1**). We also found that the method was robust and highly reproducible as evidenced by the high concordance between the three individual experiment replicates (**Supplementary Fig. 9** and **Supplementary Dataset 2**).

$KRAS^{G12V}$ and $MEK1^{DD}$ abundance effect on signaling

We first studied a well-known signaling circuit. Constitutively active mutants of KRAS and MEK1 ($KRAS^{G12V}$ and $MEK1^{DD}$) lead to ERK phosphorylation and activate components downstream in the MAPK/ERK pathway. As expected, we found that overexpression

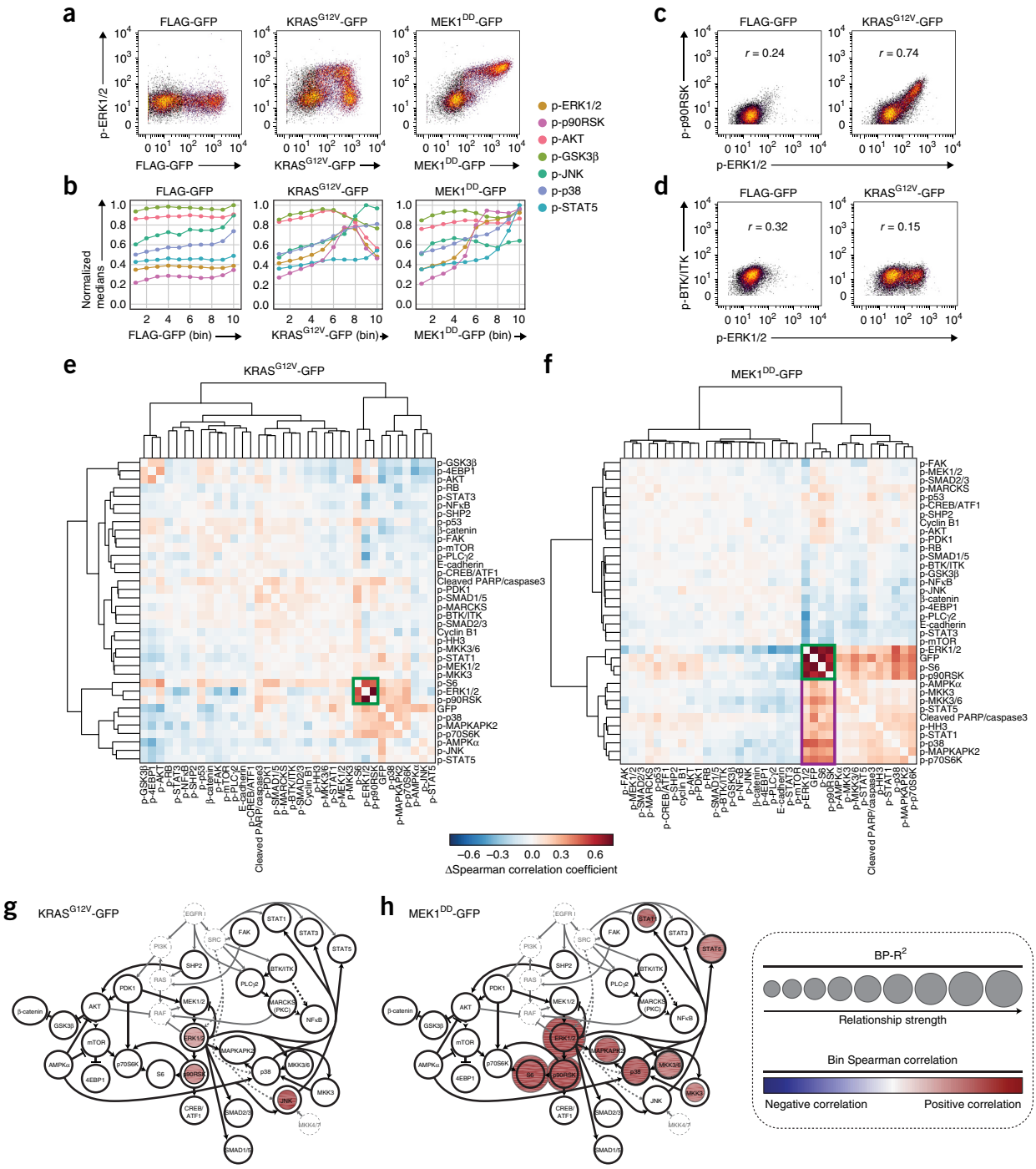

Figure 2 MAPK/ERK pathway mutants induce oncogenic signaling. (**a**) Biaxial plots of GFP, representing the abundance of the overexpressed mutant POIs, versus abundance of phosphorylation on Thr202/Tyr204 on ERK1/2. Constitutively active KRAS^{G12V}-GFP shows an absence of induction on Thr202/Tyr204 on ERK1/2 at the highest levels of KRAS^{G12V}-GFP. Constitutively active MEK1^{DD}-GFP directly phosphorylates Thr202/Tyr204 on ERK1/2, and the abundance of the POI-GFP is correlated with the amount of ERK1/2 phosphorylated at these sites. The FLAG-GFP control does not affect ERK phosphorylation sites. (**b**) The abundances of measured phosphorylation sites are plotted over the range of the KRAS^{G12V}-GFP and MEK1^{DD}-GFP expression. Phosphorylation sites of the same pathway (e.g., on ERK1/2 and p90RSK, AKT and GSK3β, or p38 and JNK) show similar trends. An individual experiment is shown here. Plots for three replicates are shown in **Supplementary Figure 9b–e**. (**c**) Strong single-cell correlations within biaxial plots indicate co-regulated phosphorylation sites. (**d**) Unchanged and reduced correlations indicate unrelated phosphorylation sites. For **c** and **d**, a representative individual experiment from three replicates is shown. (**e**,**f**) Heat maps showing for all pairs of measured markers the change in Fisher-transformed Spearman correlation values for overexpression of KRAS^{G12V}-GFP (**e**) and MEK1^{DD}-GFP (**f**) when compared to the FLAG-GFP overexpression control. (**g**,**h**) BP-R^2 scores and Spearman correlations of bin medians for all measured markers in cells where KRAS^{G12V}-GFP (**g**) or MEK1^{DD}-GFP (**h**) was overexpressed overlaid on a literature-based graph of canonical signaling pathways[14,15,21,23,35,44–48]. Strong relationships identified from the BP-R^2 analysis are plotted on the signaling maps as colored circles. The sizes of circles indicate relationship strengths quantified by BP-R^2. The directionalities of relationships, as judged by Spearman correlation of bin medians, are shown by the color of the circles (positive correlation indicates that cells show generally increasing marker levels, and a negative correlation indicates decreasing marker levels as POI-GFP levels increase). For **e** to **h**, data from three individual experiment replicates were used.

of KRAS^{G12V}-GFP or MEK1^{DD}-GFP increased phosphorylation on Thr202 and Tyr204 of ERK1/2 (**Fig. 2a**). Our approach also elucidated the abundance-dependent effects on these signaling relationships. The relationship between KRAS^{G12V}-GFP and p-ERK1/2 was bow-like, as high levels of KRAS^{G12V}-GFP corresponded to reduced phosphorylation of ERK1/2. By contrast, the MEK1^{DD}-GFP abundance relationship with p-ERK1/2 was monotonic as p-ERK1/2 increased with MEK1^{DD}-GFP expression (**Fig. 2a**). These results verified the oncogenic activation of p-ERK1/2 induced by KRAS^{G12V} and MEK1^{DD}.

Next, we analyzed the impact of KRAS^{G12V}-GFP and MEK1^{DD}-GFP abundance on all measured phosphorylation sites. We divided the measured cells into ten bins according to the GFP signals and plotted the bin medians (**Fig. 2b** and **Supplementary Fig. 9b–e**). This analysis revealed that the phosphorylation site abundances on ERK1/2 and its direct downstream target Ser380 of p90RSK had similar relationships to the abundances of KRAS^{G12V}-GFP and MEK1^{DD}-GFP. Phosphorylation of AKT on Ser473 and its direct target Ser9 of GSK3β also had parallel trends and showed reduced levels when the MAPK/ERK signal peaked, suggesting interpathway regulation. We also observed increased JNK phosphorylation on Thr183/Tyr185 induced by the KRAS^{G12V} mutant (**Fig. 2b**), as reported previously[32]. This shows that our approach recapitulates known signaling relationships and identifies abundance-determined signaling responses.

We then systematically evaluated signaling relationships between all pairs of measured markers modulated by KRAS^{G12V}-GFP or MEK1^{DD}-GFP overexpression. We exploited the fact that overexpression of one protein increases signaling (i.e., phosphorylation levels) and thus expands the dynamic range of many measured markers (**Fig. 2c**). This enabled the use of correlation analysis to distinguish signaling relationships (high correlation) from biological and technical noise (low correlation). For example, overexpression of KRAS^{G12V}-GFP resulted in an increased Spearman correlation between p-ERK1/2 and p-p90RSK compared to control (**Fig. 2c**), whereas ERK-independent phosphorylation sites, such as Tyr551 of BTK/ITK, showed low correlation with p-ERK1/2 levels in both control and overexpression conditions (**Fig. 2d**).

Identifying changes in pairwise Spearman correlations for all measured markers in the KRAS^{G12V}-GFP and MEK1^{DD}-GFP overexpression data compared to the FLAG-GFP control enabled systematic analysis of signaling relationship patterns (**Fig. 2e**,**f**). Phosphorylation levels of proteins in the MAPK/ERK pathways showed strong increases in correlation, and pathway members clustered together (**Fig. 2e**,**f**, green squares). We also observed that phosphorylation of MAPK/p38 pathway members and STAT proteins (STAT1 and STAT5) was increasingly correlated with levels of MAPK/ERK pathway members as MEK^{DD}-GFP levels increased (**Fig. 2f**, purple rectangle), indicating crosstalk between MAPK and STAT pathways. These results reveal relationships among many measured markers and show that increases in correlation reflect pathways and grouped biological processes.

Automated analysis of abundance-induced signaling

Spearman correlation analysis can uncover strictly monotonic relationships between phosphorylation levels on signaling proteins; however, protein-abundance-dependent signaling responses can be complex (**Fig. 2a**, see KRAS^{G12V}). We therefore developed a density-independent measure BP-R^2 to quantify the strengths of relationships between the abundance of a POI and measured phosphorylation sites. BP-R^2 creates ten bins across the POI-GFP expression range and calculates the relationship strength considering bin medians and the global mean (**Supplementary Fig. 10a**,**b**, Online Methods, and **Supplementary Software**). Using the BP-R^2 values for all negative controls, a cutoff for strong signaling relationships was determined (**Supplementary Fig. 10c**). Benchmarking BP-R^2 in identifying strong signaling relationships from the overexpression data sets showed that BP-R^2 outperformed methods often used for this task[30,33] (**Supplementary Fig. 11a**,**b**). The strong relationships identified by BP-R^2 were plotted in a two-dimensional layout guided by canonical pathways (**Fig. 2g**,**h**). The directionality of measured signaling relationships was determined by Spearman correlation of the bin medians (**Supplementary Fig. 10b**). A positive correlation indicates that cells show generally increasing marker levels, and a negative correlation indicates generally decreasing marker levels as POI-GFP levels increase.

Analysis of KRAS^{G12V}-GFP and MEK1^{DD}-GFP overexpression versus all measured markers using BP-R^2 revealed strong, positively correlated relationships of MEK^{DD}-GFP to downstream MAPK/ERK pathway nodes. KRAS^{G12V}-GFP levels, although also positively correlated with MAPK/ERK nodes, exhibited the same, but weaker relationships (**Fig. 2a**,**b**,**g**,**h**). Together, these results suggest that feedback regulation of upstream MAPK nodes differs between the studied mutants. Additionally, this network view revealed that MEK1^{DD}-GFP abundance had a strong positive impact on nodes in the MAPK/p38 pathway; the previously observed KRAS^{G12V}-induced phosphorylation of JNK[32] was dependent on KRAS^{G12V} abundance (**Fig. 2g**,**h**). These results show that overexpression of signaling proteins, in conjunction with BP-R^2 and correlation analysis, identifies known relationships and is a valid platform for discovery of signaling relationships in a comprehensive and abundance-dependent manner.

Table 1 Overexpressed signaling proteins

Overexpressed proteins	Gene ID	UniProt entry
SRC	*SRC*	P12931
PDK1	*PDPK1*	O15530
AKT1	*AKT1*	P31749
GSK3β	*GSK3B*	P49841
MKK7	*MAP2K7*	O14733
MKK6	*MAP2K6*	P52564
p38α	*MAPK14*	Q16539
ERK2	*MAPK1*	P28482
p90RSK	*RPS6KA1*	Q15418
CRAF	*RAF1*	P04049
JNK1	*MAPK8*	P45983
p110α	*PIK3CA*	P42336
BRAF	*BRAF*	P15056
ASK1	*MAP3K5*	Q99683
p70S6K	*RPS6KB1*	P23443
MEK1	*MAP2K1*	Q02750
KRAS	*KRAS*	P01116
HRAS	*HRAS*	P01112
SHP2	*PTPN11*	Q06124
S6	*RPS6*	P62753

Node abundance dependency analyses of the EGFR network

To study the node abundance dependency of signaling relationships and dynamics in the EGFR signaling network, we overexpressed 20 EGFR-related signaling proteins individually in HEK293T cells (**Table 1**). Each of the 20 GFP-tagged POIs was validated in previous studies (**Supplementary Table 5**) and in our system (**Supplementary Figs. 2** and **3** and **Supplementary Dataset 1**). 18 h after transfection, we treated cells with EGF and quantified signaling by mass cytometry over a 60-min time course. To exclude signaling relationships caused by channel-to-channel spillover, we applied a stringent experimental filter (**Supplementary Fig. 12**). The median marker intensities during the time course are shown in **Supplementary Figure 13a**. Based on these data we performed two sets of analyses. In the first, we used

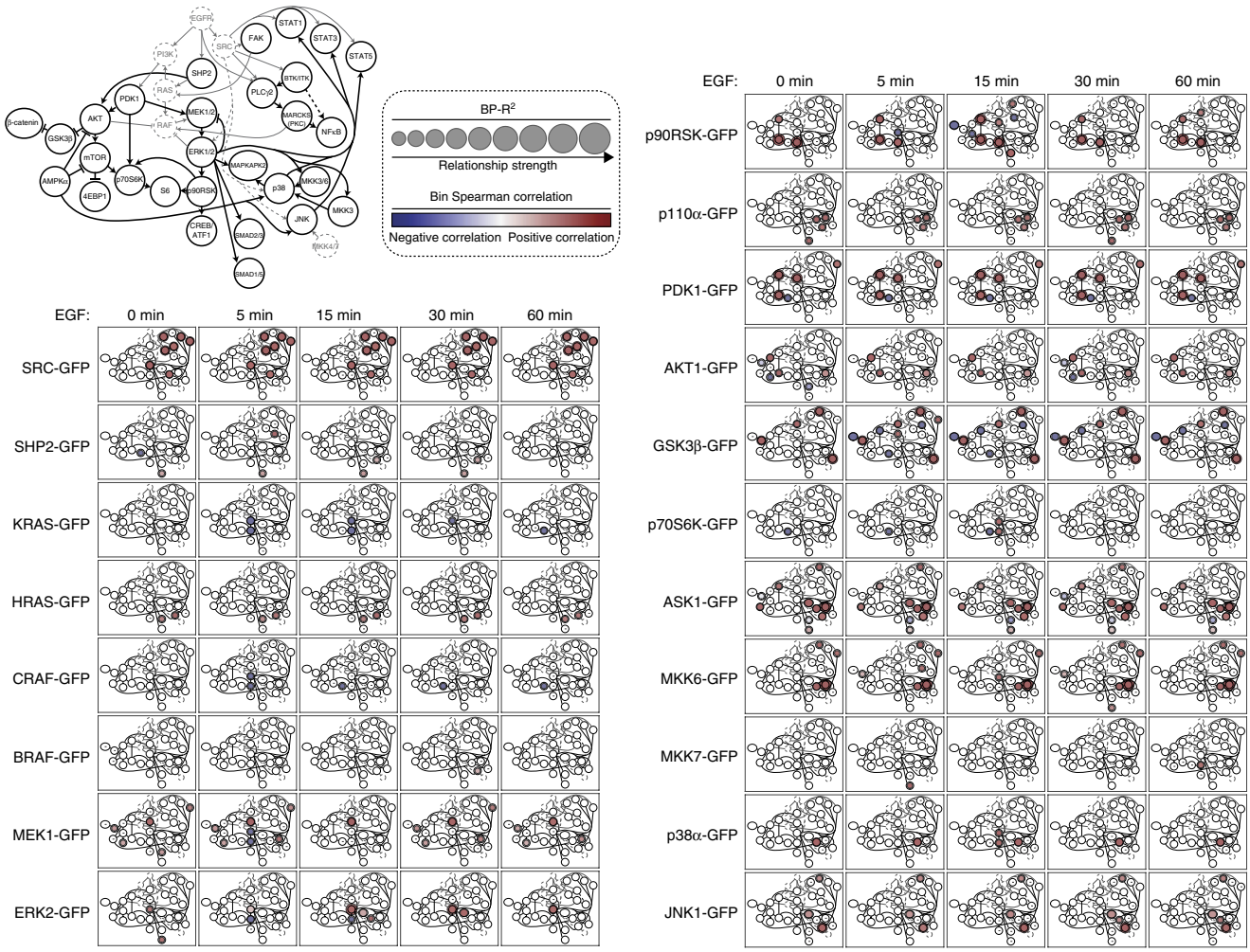

Figure 3 Analysis of abundance-dependent dynamics of EGFR signaling. HEK293T cells overexpressing GFP-tagged signaling proteins listed in **Table 1** were treated with EGF for 0, 5, 15, 30, and 60 min. Strong abundance-dependent signaling relationships (**Supplementary Fig. 10c**) are plotted on the signaling map with circle sizes and colors indicating strengths (BP-R^2 score) and directionalities (Spearman correlation of bin medians), respectively. The miniaturized network is the same as used in **Figure 2**. Overexpression of S6-GFP did not induce any strong signaling relationships (data not shown). For all analyses, data from three individual experiment replicates were used.

BP-R^2 analysis and Spearman correlations to evaluate how the abundance of overexpressed proteins influenced phosphorylation at the measured sites (**Fig. 3**, **Supplementary Fig. 13b**, and **Supplementary Datasets 2–4**). In the second, we examined how features of signaling dynamics depend on protein abundance (**Fig. 4**).

In the first analysis, strong and broad signaling responses to overexpression were identified for the upstream kinases PDK1-, GSK3β-, SRC-, and ASK1-GFP without EGF stimulation (**Fig. 3**). Overall, we identified 59 strong signaling relationships in the unstimulated conditions. Overexpression of many kinases induced strong and positively correlated signaling relationships with their own phosphorylation (**Fig. 3** and **Supplementary Dataset 4**). Overexpression of CRAF-, KRAS-, p70S6K-GFP, and others induced signaling responses only upon EGF stimulation (**Fig. 3**). Notably, under stimulated conditions, KRAS-, CRAF-, and MEK1-GFP levels negatively correlated with phosphorylation levels of downstream kinases p-ERK1/2 and p-p90RSK (**Fig. 3**). Activating mutations in KRAS and CRAF (**Fig. 2**), but not protein overexpression alone, may activate oncogenic signaling.

To systematically assess signaling relationships identified by BP-R^2, we used the literature-based signaling network, SIGNOR[34]. For each relationship, we computed the shortest signed directed path length according to the SIGNOR network (**Supplementary Table 6**). We found that 76% of the strong relationships identified in the unstimulated conditions had paths with a maximum of three steps, highlighting that our approach identifies rather direct signaling relationships. Only 14 abundance-dependent relationships with four or more path steps were identified. Comparison of our strong signaling relationships with literature indicated that many EGF signaling connections that we identified were previously reported. We also propose many relationships that have—to our knowledge—not been previously reported, for example: p90RSK to PDK1 (Ser241), GSK3β to SHP2 (Tyr580), JNK1 to MAPKAPK2 (Thr334), p110α to MKK3 (Ser189), p110α to MKK6 (Ser207), ASK1 to PDK1 (Ser241), ASK1 to GSK3β (Ser9), and ASK1 to AMPKα (Thr172) (**Table 2**).

Phosphorylation levels of many members of the MAPK/ERK pathways showed complex relationships (i.e., measured phosphorylation levels varied over the analyzed POI-GFP range and the relationships

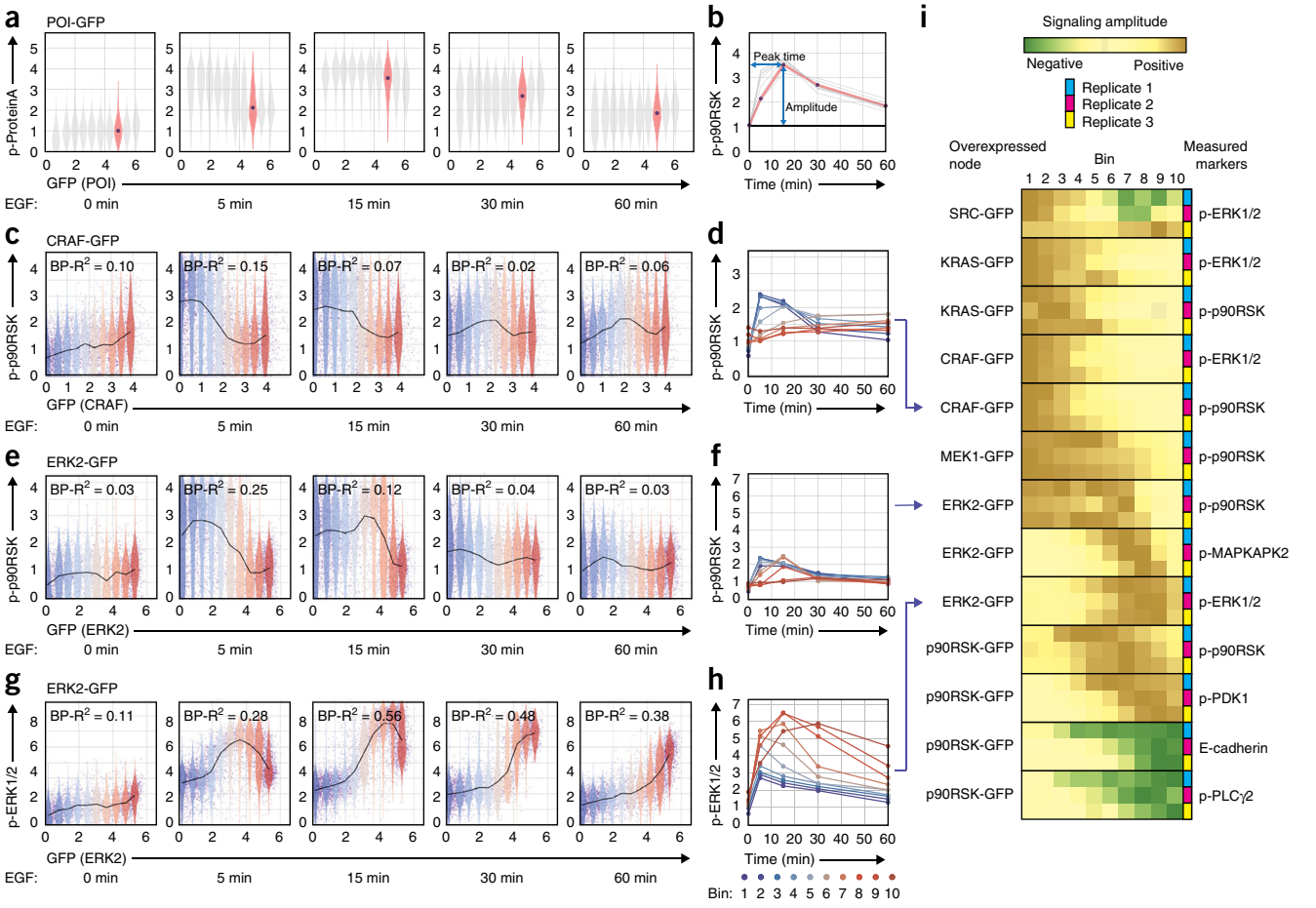

Figure 4 Analysis of node abundance-dependent EGFR signaling dynamics. (**a**,**b**) Schematic plots of amplitude and peak-time analysis. (**a**) The *x* axis (i.e., overexpressed protein as determined by the GFP measurement) was split into ten bins. (**b**) Median phosphorylation abundance in each bin was plotted on the *y* axis versus time (*x* axis) to visualize abundance dependency of signaling dynamics. (**c**,**d**) Mass cytometry ion counts (arcsinh transformed) measured for p-p90RSK (*y* axis) as a function of ion counts measured for abundance of CRAF-GFP (*x* axis) and EGF stimulation time. (**e**–**h**) The same layouts for ERK2-GFP abundance-determined p-p90RSK levels (**e**,**f**) and p-ERK1/2 levels (**g**,**h**) are shown. (**i**) Heat map showing protein abundances with strong influences on signaling amplitudes with color indicating normalized signaling amplitudes. Only overexpressed proteins with an amplitude-ratio higher than threefold for more than two of the three replicates were identified as strong influences and are included in the heat map. For **a**–**h**, a representative individual experiment from the three replicates is shown. Other replicates are presented in **Supplementary Dataset 5**. In **i**, all replicate data are shown.

did not fit linear, sigmoidal, or quadratic models) with levels of POI-GFPs upon EGF stimulation. These relationships can be explained by abundance-dependent modulation of the signaling dynamics in response to EGF. Thus, in the second set of analyses we examined how signaling dynamics, as quantified by amplitude and peak time (maximum amplitude for each bin), depended on abundance of an overexpressed protein (**Fig. 4**). In order to view signaling trajectories as functions of protein abundance, we binned the POI-GFP levels into ten bins (**Fig. 4a** and **Supplementary Dataset 2**). This allowed us to trace the signaling trajectories of cells with similar protein overexpression levels (i.e., those in the same bin) over the EGF stimulation time course (**Fig. 4b** and **Supplementary Dataset 5**). Strong and robust changes in signaling amplitudes (**Fig. 4c**–**i**) and peak times (**Supplementary Fig. 14**) were found. Notably, the maximum amplitudes were independent of the overexpression range of a given POI (**Supplementary Fig. 15**).

We found that high CRAF-GFP and KRAS-GFP abundance strongly reduced signaling amplitudes of p-ERK1/2 and p-p90RSK (**Fig. 4c**,**d**,**i**), whereas high abundance of MEK1-GFP strongly reduced amplitudes and delayed peak times for p-p90RSK (**Fig. 4i** and **Supplementary Fig. 14**). Overexpression of ERK2-GFP led to complex abundance-dependent responses of p-p90RSK and p-ERK1/2 after EGF stimulation (**Fig. 4e**–**h**). p-ERK1/2 amplitudes increased and peak times were delayed as a function of ERK2-GFP abundance level (**Fig. 4g**–**i** and **Supplementary Fig. 14**). Intermediate abundance levels of ERK2-GFP also delayed the p-p90RSK peak times relative to low ERK2-GFP abundance, whereas cells with high ERK2-GFP levels exhibited minimal p-p90RSK signaling dynamics (**Fig. 4e**,**f**,**i** and **Supplementary Fig. 14**). Overexpression of p90RSK-GFP modulated the signaling amplitude of its potential crosstalk phosphorylation site, Ser241 of PDK1, in an abundance-dependent manner, and increasing expression of p90RSK increased p-PDK1 amplitudes (**Fig. 4i**). Thus, we observed abundance-dependent signaling dynamics across the range of overexpression levels. Overexpression of upstream signaling proteins (KRAS-, CRAF-, MEK1-, and ERK2-GFP) in the MAPK/ERK pathway led to reduced signaling amplitudes and delayed peak times of their downstream targets. These observations show that our approach can quantify the role of protein abundance in determining the dynamic signaling response to an extracellular stimulation.

Table 2 Relationships with shortest signed directed path length above 3 in the SIGNOR database

Overexpressed POI	Target	Sign	Shortest signed directed path (SIGNOR)	Literature information
SRC	p-BTK/ITK	1	6	SRC family kinases phosphorylate BTK[48]
SHP2	p-S6	–1	5	Known regulation[49]
ASK1	p-PDK1	1	5	Potential novel relationship
SRC	p-PLCγ2	1	5	SRC family kinases activates PLCγ2 (ref. 48)
ASK1	p-AMPKα	1	4	Potential novel relationship
GSK3β	p-SHP2	1	4	Potential novel relationship
p90RSK	p-PDK1	1	4	Potential novel relationship
JNK1	p-STAT1	1	4	JNK activates STAT1 (ref. 50)
JNK1	p-MAPKAPK2	1	4	Potential novel relationship
p110α	p-MKK3/6	1	4	Potential novel relationships
HRAS	p-SMAD2/3	1	4	Known crosstalk[21]
ASK1	p-GSK3β	1	4	Potential novel relationships
PDK1	p-S6	–1	4	Overexpression-induced negative regulation
p70S6K	p-S6	–1	4	Overexpression-induced negative regulation

DISCUSSION

Here, we present an approach coupling transient overexpression with mass-cytometry-based single-cell measurements to characterize signaling-network activation states and signaling dynamics over a quasi-continuous, high dynamic range of protein abundance. To highlight the utility of our approach, we present a comprehensive single-cell proteomic analysis of the EGFR network that enabled an analysis of abundance-dependent effects of signaling proteins on the state and dynamics of the signaling network. We evaluated the effects of overexpressing 20 EGFR network key nodes with a 60-min EGF stimulation time course. In each of the 360 conditions, we measured the effect of a POI over an abundance range of four orders of magnitude on 35 markers by mass cytometry providing a unique and valuable quantitative single-cell resource of abundance dependencies of EGFR signaling.

Previously, the heterogeneity of protein levels after transient transfection was considered problematic. Here, we took advantage of this cell-to-cell variation as it results in a continuous titration of protein abundance over four orders of magnitude. Untransfected cells also provided an internal control for each experiment. We used the multiplexing capabilities of mass cytometry to characterize abundance dependencies of signaling network state and dynamics. Applied to the EGFR signaling network, our approach recapitulated known relationships, suggested ones previously not described, and revealed the intricate modulation of signal amplitudes and peak times as functions of continuous protein abundance.

Our approach contributes to the understanding of signaling on several levels. First, the approach can be used to study uncharacterized proteins and to suggest additional roles for characterized ones. Second, we were able to directly relate POI abundance with the comprehensive analysis of signaling dynamics in response to stimulation. Such analyses are necessary for an understanding of differential signal processing in identical cell types and in disease states characterized by heterogeneity in protein expression such as cancer. Third, the overexpression yields a large dynamic range of signaling activity and can reveal signaling relationships masked by stochastic processes and technical noise under otherwise similar conditions, facilitating the computational analysis of signaling relationships. Fourth, we present a metric termed BP-R^2, which allows the quantification of the strengths of arbitrarily shaped signaling relationships. BP-R^2 was superior to state-of-the-art methods for analysis of our data set. Fifth, and finally, we were able to infer protein-abundance-dependent signaling kinetics from single-cell snapshot data.

Our approach recapitulated known oncogenic signaling behaviors induced by the constitutively active mutants $KRAS^{G12V}$ and $MEK1^{DD}$ and identified novel abundance-dependent signaling relationships. For example, p-ERK1/2 was attenuated in cells with highly overexpressed $KRAS^{G12V}$-GFP, potentially due to negative feedback loops or senescence[35]. Overexpression of the wild-type KRAS-GFP and MEK1-GFP did not induce downstream signaling activation, suggesting that mutations on KRAS or MEK1 are the main drivers of oncogenic signaling. Further, our approach allows study of abundance-dependent signaling dynamics. In the MAPK/ERK pathway, high abundance of upstream signaling mediators KRAS, CRAF, MEK1, or ERK2 reduced amplitudes and delayed peak times of downstream phosphorylation sites. One possible explanation is that the signal transduction is determined by the competition between active and inactive forms of a signaling protein for substrates. Overexpression increases the total abundance but may reduce the percentage of the active form.

KRAS amplification has been identified in many cancer types. Amplification, however, is not correlated with the phosphorylation of ERK1/2 (ref. 36). Rather, *KRAS* amplification mediates resistance to inhibitors targeting growth-pathway-related kinases, including EGFR, MET, and MEK1/2; KRAS knockdown diminishes drug resistance[37–39]. Our results indicate that owing to reduced downstream signaling amplitudes in response to EGF stimulation, the dependency of cells on the MAPK/ERK pathway may decrease upon KRAS overexpression, suggesting a mechanism for cancer cell resistance to inhibitors.

Comparing the strong signaling relationships we identified with those in the SIGNOR database, we propose signaling relationships not described previously. (1) Our data suggest that p90RSK potentially forms a positive feedback loop and activates the upstream signaling protein PDK1. (2) GSK3β has been identified as a central signaling controller and has multiple substrates[40]; our results suggest that SHP2 is a potential direct or indirect target of GSK3β. (3) We also propose that JNK1 is a MAPKAPK2 activator. (4) PI3K and MKK3/6 are known to be regulated by RAC1 (ref. 41); our results suggest PI3K activates MKK3/6 independently. (5) Recent studies indicate that ASK1 contributes to negative regulation of PDK1 through phosphorylation on Thr254 of PDK1 (ref. 42). We observed ASK1 overexpression-induced PDK1 phosphorylation on Ser241, inducing PDK1 activity and downstream GSK3β phosphorylation on Ser9. (6) In addition to the known AMPK-mediated ASK1 activation[43], our data indicate ASK1 activation of AMPKα via phosphorylation on Thr172. (7) We have also observed negative correlations between the abundance of p70S6K or PDK1 to the phosphorylation level of S6 (Ser235/Ser236), indicating overexpression-induced-negative-feedback regulation.

Our method has several limitations. First, we do not measure the endogenous expression level of the POI. However, exogenous expression is linearly correlated with the total protein level (**Supplementary Fig. 4a**), validating GFP as readout of the total POI. Second, all results

in mass cytometry rely on antibodies; for this work, all antibodies were thoroughly validated (**Supplementary Table 3**). Third, we do not measure the abundance range of the studied proteins in cancer cells, however, proteome studies of cancer cells and databases such as PaxDb[11] indicate a range similar to those studied here. Fourth, high expression levels of a protein kinase may induce non-specific phosphorylation; however, our data allow choosing the analyzed expression range *in silico*, thus such effects can be excluded.

The approach described here provides a method to study how the abundance variance of signaling proteins in different tissues and cell lines results in distinct signaling behaviors. The application of our approach to synthetic biology, stem cell biology, developmental biology, and cancer-related processes, such as the epithelial-mesenchymal transition, will enable quantitative identification of key proteins and signaling determinants in cell differentiation at phenotypical switching points. We envision that determining which signaling relationships and thresholds enable diseased cells to overcome drug treatment will be a highly relevant application.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank the Bodenmiller laboratory for support and fruitful discussions, the Lehner laboratory and the Mosimann laboratory for sharing equipment. We would especially like to thank A.-C. Gingras, Lunenfeld-Tanenbaum Research Institute, for sharing the pDEST vectors used in this study. This work was supported by the Swiss National Science Foundation (SNSF) R'Equip grant 316030-139220, a SNSF Assistant Professorship grant PP00P3-144874, a Swiss Cancer League grant, the PhosphonetPPM SystemsX grant, and funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 336921. The work of J.D.W. was supported by a National Science Foundation Graduate Research Fellowship under grant no. DGE-1650044 and a Whitaker International Fellowship awarded by the Institute of International Education. The work of D.S. is supported by the Forschungskredit of the University of Zurich Fellowship under grant no. FK-74419-01-01.

AUTHOR CONTRIBUTIONS

X.-K.L. and B.B. conceived and designed the experiments. X.-K.L., M.T. and N.D. performed experiments. V.R.T.Z., J.D.W., X.-K.L. and D.S. performed data analysis. X.-K.L. and B.B. wrote the manuscript. All authors commented on and edited the final version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Wolf-Yadlin, A. *et al.* Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* **2**, 54 (2006).
2. De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).
3. Feinberg, A.P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).
4. Bywater, M.J., Pearson, R.B., McArthur, G.A. & Hannan, R.D. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat. Rev. Cancer* **13**, 299–314 (2013).
5. Silvera, D., Formenti, S.C. & Schneider, R.J. Translational control in cancer. *Nat. Rev. Cancer* **10**, 254–266 (2010).
6. Santarius, T., Shipley, J., Brewer, D., Stratton, M.R. & Cooper, C.S. A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **10**, 59–64 (2010).
7. Govindarajan, B. *et al.* Overexpression of Akt converts radial growth melanoma to vertical growth melanoma. *J. Clin. Invest.* **117**, 719–729 (2007).
8. Eralp, Y. *et al.* MAPK overexpression is associated with anthracycline resistance and increased risk for recurrence in patients with triple-negative breast cancer. *Ann. Oncol.* **19**, 669–674 (2008).
9. Han, T. *et al.* PTPN11/Shp2 overexpression enhances liver cancer progression and predicts poor prognosis of patients. *J. Hepatol.* **63**, 651–660 (2015).
10. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
11. Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).
12. Citri, A. & Yarden, Y. EGF-ERBB signalling: towards the systems level. *Nat. Rev. Mol. Cell Biol.* **7**, 505–516 (2006).
13. Tebbutt, N., Pedersen, M.W. & Johns, T.G. Targeting the ERBB family in cancer: couples therapy. *Nat. Rev. Cancer* **13**, 663–673 (2013).
14. Roberts, P.J. & Der, C.J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310 (2007).
15. Mendoza, M.C., Er, E.E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36**, 320–328 (2011).
16. Olayioye, M.A., Neve, R.M., Lane, H.A. & Hynes, N.E. The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* **19**, 3159–3167 (2000).
17. Manning, B.D. & Cantley, L.C. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261–1274 (2007).
18. Bowman, T., Garcia, R., Turkson, J. & Jove, R. STATs in oncogenesis. *Oncogene* **19**, 2474–2488 (2000).
19. Oliva, J.L., Griner, E.M. & Kazanietz, M.G. PKC isozymes and diacylglycerol-regulated proteins as effectors of growth factor receptors. *Growth Factors* **23**, 245–252 (2005).
20. Kim, D., Rath, O., Kolch, W. & Cho, K.-H. A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. *Oncogene* **26**, 4571–4579 (2007).
21. Massague, J. Integration of Smad and MAPK pathways: a link and a linker revisited. *Genes Dev.* **17**, 2993–2997 (2003).
22. Zhang, Y. *et al.* Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250 (2005).
23. Kim, S.Y. *et al.* AMP-activated protein kinase-α1 as an activating kinase of TGF-β-activated kinase 1 has a key role in inflammatory signals. *Cell Death Dis.* **3**, e357 (2012).
24. Corcoran, R.B. *et al.* Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models. *Cancer Cell* **23**, 121–128 (2013).
25. Tewari, M. *et al.* Systematic interactome mapping and genetic perturbation analysis of a *C. elegans* TGF-β signaling network. *Mol. Cell* **13**, 469–482 (2004).
26. Sundqvist, A. *et al.* Specific interactions between Smad proteins and AP-1 components determine TGFβ-induced breast cancer cell invasion. *Oncogene* **32**, 3606–3615 (2013).
27. Aoki, K. *et al.* Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Mol. Cell* **52**, 529–540 (2013).
28. Bodenmiller, B. *et al.* Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* **30**, 858–867 (2012).
29. Bendall, S.C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
30. Krishnaswamy, S. *et al.* Systems biology. Conditional density-based analysis of T cell signaling in single-cell data. *Science* **346**, 1250689 (2014).
31. Couzens, A.L. *et al.* Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci. Signal.* **6**, rs15 (2013).
32. Zhou, Y. *et al.* Chimeric mouse tumor models reveal differences in pathway activation between ERBB family- and KRAS-dependent lung adenocarcinomas. *Nat. Biotechnol.* **28**, 71–78 (2010).
33. Redell, M.S. *et al.* FACS analysis of Stat3/5 signaling reveals sensitivity to G-CSF and IL-6 as a significant prognostic factor in pediatric AML: a Children's Oncology Group report. *Blood* **121**, 1083–1093 (2013).
34. Perfetto, L. *et al.* SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* **44**, D548–D554 (2016).
35. Xu, Y., Li, N., Xiang, R. & Sun, P. Emerging roles of the p38 MAPK and PI3K/AKT/mTOR pathways in oncogene-induced senescence. *Trends Biochem. Sci.* **39**, 268–276 (2014).
36. Rahman, M.T. *et al.* KRAS and MAPK1 gene amplification in type II ovarian carcinomas. *Int. J. Mol. Sci.* **14**, 13748–13762 (2013).
37. Valtorta, E. *et al.* KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy. *Int. J. Cancer* **133**, 1259–1265 (2013).
38. Cepero, V. *et al.* MET and KRAS gene amplification mediates acquired resistance to MET tyrosine kinase inhibitors. *Cancer Res.* **70**, 7580–7590 (2010).
39. Little, A.S. *et al.* Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. *Sci. Signal.* **4**, ra17 (2011).
40. Cohen, P. & Frame, S. The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2**, 769–776 (2001).

41. Shin, I., Kim, S., Song, H., Kim, H.-R.C. & Moon, A. H-Ras-specific activation of Rac-MKK3/6-p38 pathway: its critical role in invasion and migration of breast epithelial cells. *J. Biol. Chem.* **280**, 14675–14683 (2005).
42. Seong, H.-A., Jung, H., Ichijo, H. & Ha, H. Reciprocal negative regulation of PDK1 and ASK1 signaling by direct interaction and phosphorylation. *J. Biol. Chem.* **285**, 2397–2414 (2010).
43. Lee, Y.-K., Hwang, J.-T., Kwon, D.Y., Surh, Y.-J. & Park, O.J. Induction of apoptosis by quercetin is mediated through AMPKalpha1/ASK1/p38 pathway. *Cancer Lett.* **292**, 228–236 (2010).
44. Cardaci, S., Filomeni, G. & Ciriolo, M.R. Redox implications of AMPK-mediated signal transduction beyond energetic clues. *J. Cell Sci.* **125**, 2115–2125 (2012).
45. Rawlings, J.S., Rosler, K.M. & Harrison, D.A. The JAK/STAT signaling pathway. *J. Cell Sci.* **117**, 1281–1283 (2004).
46. Nyati, M.K., Morgan, M.A., Feng, F.Y. & Lawrence, T.S. Integration of EGFR inhibitors with radiochemotherapy. *Nat. Rev. Cancer* **6**, 876–885 (2006).
47. Mitra, S.K., Hanson, D.A. & Schlaepfer, D.D. Focal adhesion kinase: in command and control of cell motility. *Nat. Rev. Mol. Cell Biol.* **6**, 56–68 (2005).
48. Hendriks, R.W., Yuvaraj, S. & Kil, L.P. Targeting Bruton's tyrosine kinase in B cell malignancies. *Nat. Rev. Cancer* **14**, 219–232 (2014).
49. Marin, T.M. *et al.* Shp2 negatively regulates growth in cardiomyocytes by controlling focal adhesion kinase/Src and mTOR pathways. *Circ. Res.* **103**, 813–824 (2008).
50. Wei, Z.Z. *et al.* Regulatory role of the JNK-STAT1/3 signaling in neuronal differentiation of cultured mouse embryonic stem cells. *Cell. Mol. Neurobiol.* **34**, 881–893 (2014).

ONLINE METHODS

Cloning. DNA sequences of the genes of interest were provided in entry clones by William Hahn and David Root[51] (via Addgene and NEXUS Personalized Health Technologies at ETH Zurich). Destination vectors, including pDEST pcDNA5 FRT TO-eGFP, pDEST 5′ Triple Flag pCDNA5 FRT TO and pDEST 3′ Triple Flag pcDNA5 FRT TO, were kindly provided by Anne-Claude Gingras at Lunenfeld-Tanenbaum Research Institute, Toronto, Canada[31]. Expression vectors encoding the FLAG- or GFP-tagged fusion proteins were generated via Gateway Cloning and sequenced before transfection.

Cell culture. HEK293T cells, obtained from ATCC, were cultured in DMEM (D5671, SIGMA), supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. For cell passaging or harvesting, cells were incubated with 1× TrypLE Express (Life Technologies) for 2 min at 37 °C. Purity and sterility of the cell line were certified by ATCC. Mycoplasma was not detected with the LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich).

Transfection and stimulation. HEK293T cells were seeded at the density of 0.7 million per well in 6-well plates. After 24 h, cells were transfected with 2 µg plasmid and 4 µl of jetPRIME (PolyPlus) per well with the standard protocol provided by the manufacturer. At 18 h after transfection, EGF (Peprotech) was added to a final concentration of 100 ng/ml. At 20 min before a given EGF stimulation time point, 5-Iodo-2-deoxyuridine (IdU) was added to the medium at the final concentration of 10 µM. At 2 min before a given EGF stimulation time point, medium was replaced by 1× TrypLE to induce cell detachment. At the time point, paraformaldehyde (PFA, from Electron Microscopy Sciences) was added to the cell suspension to a final percentage of 1.6%, and cells were incubated at room temperature for 10 min. If EGF stimulation was not necessary in the experiment, cells were directly harvested and crosslinked with PFA. Crosslinked cells were washed twice with cell staining media (CSM, PBS with 0.5% BSA, 0.02% NaN_3) and after centrifugation, ice-cold methanol was used to resuspend the cells, followed by a 10-min permeabilization on ice or for long-term storage at −80 °C. Three individual experimental replicates were performed for each experiment. In each replicate, the experimental procedures were performed on different days.

Cell sorting. HEK293T cells overexpressing FLAG-GFP were detached from the plates as described above and resuspended in the FACS buffer (PBS with 0.5% BSA and 5 mM EDTA). Cells were sorted with BD FACSAria III Cell Sorter into GFP low, intermediate, and high levels with the strategy indicated in **Supplementary Figure 5**.

Live cell imaging. HEK293T cells were seeded in CultureWell Chambered Coverglass (Thermo Fisher Scientific) pre-coated with fibronectin. Transfection of FLAG-GFP was performed as described above. At 18 h after transfection, cells were imaged with a Leica DMI 6000 inverted microscope at 37 °C and 5% CO_2. Images were taken every 2 min for 1 h.

Immunofluorescence staining. Cells were cultured in 16-well glass chamber slides (Thermo Fisher Scientific). Transfection was done as described above. Before staining, culture medium was removed, and the slide was then washed with PBS. To crosslink cells, 4% PFA was added, and cells were incubated at room temperature for 20 min. Slides were washed with PBS three times, and cells were permeabilized for 5 min with 0.1% TritonX-100 dissolved in PBS at room temperature. After washing three times with PBS, cells were incubated in blocking buffer (10% goat serum diluted in PBS) for 30 min at room temperature. Primary (anti-GFP, FM264G, BioLegend, 1:200) and secondary (Goat anti-Rat Alexa Fluor 568, 1:500, supplemented with Hoechst 33342 at a final concentration of 100 µg/ml) antibodies were diluted in blocking buffer and applied to slides. A total protein stain of Alexa Fluor 647 Succinimidyl Ester (Life Technologies) was used to indicate cell outlines. Cells were washed three times with PBS after each incubation step. Slides were mounted with ProLong Gold Antifade Reagent (Life Technologies) before imaging with a CLSM Leica TCS SP8 microscope.

Antibody conjugation. The MaxPAR antibody conjugation kit (Fluidigm) was used to generate isotope-labeled antibodies using the manufacturer's standard protocol. After conjugation, the antibody yield was determined based on absorbance of 280 nm. Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH) was used to dilute antibodies for long-term storage at 4 °C.

Barcoding and staining protocol. Formalin-crosslinked and methanol-permeabilized cells were washed three times with CSM and once with PBS. Cells were incubated in PBS containing barcoding reagents (^{102}Pd, ^{104}Pd, ^{105}Pd, ^{106}Pd, ^{108}Pd, ^{110}Pd, ^{113}In and ^{115}In) at a final concentration of 100 nM for 30 min at room temperature and then washed three times with CSM[28]. Barcoded cells were then pooled and stained with the metal-conjugated antibody mix (**Supplementary Dataset 1**) at room temperature for 1 h. The antibody mix was removed by washing cells three times with CSM and once with PBS. For DNA staining, iridium-containing intercalator (Fluidigm) diluted in PBS with 1.6% PFA was incubated with the cells at 4 °C overnight. On the day of the measurement, the intercalator solution was removed, and cells were washed with CSM, PBS, and ddH_2O. After the last washing step, cells were resuspended in ddH_2O and filtered through a 70-µm strainer.

Mass cytometry analysis. EQ Four Element Calibration Beads (Fluidigm) were added to cell suspensions in a 1:10 ratio (v/v). Samples were analyzed on a CyTOF2 (Fluidigm). The manufacturer's standard operation procedures were used for acquisition at a cell rate of ~500 cells per second. After the acquisition, all FCS files from the same barcoded sample were concatenated[28]. Data were then normalized, and bead events were removed[52] before doublet removal and de-barcoding of cells into their corresponding wells using a doublet-filtering scheme and single-cell deconvolution algorithm[53]. Subsequently, data were processed using Cytobank (http://www.cytobank.org/). Additional gating on the DNA channels (^{191}Ir and ^{193}Ir) and ^{139}La/^{141}Pr was used to remove remaining doublets, debris and contaminating particulates.

Data visualization and analysis.

Bi-axis scatter plots. Bi-axis scatter plots were generated in Cytobank (http://www.cytobank.org/).

Data preprocessing. Raw data were transformed using the inverse hyperbolic sine transform with a cofactor of 5:

$$data = arcsinh(data_{raw} / 5)$$

Except where use of raw data values is specifically noted, all visualizations and analyses were performed using transformed data.

Data binning. For data binning, the range between the lower and upper 2.5% of observations was divided into ten equal bins: bin_1,.,bin_{10}. The observations in the lower and upper 2.5% were assigned to the lowest and highest bins, respectively. In order to be able to compare expression levels between samples within a time course replicate, all observations of the time course were used to determine the binning.

Correlation analysis. Spearman correlation (r_{ij}) was calculated between all marker pairs (i, j) for each replicate and condition. Fisher's z-transformation:

$$z_{i,j} = arctanh(r_{ij})$$

was used to compare pairwise correlation coefficients across conditions. For each overexpressed POI, the change in correlation matrix was calculated by subtracting the median $\tilde{z}_{ij}$ value (across replicates) of the FLAG-GFP controls from the median $\tilde{z}_{ij}$ value (across replicates) of the overexpression condition.

$$\Delta\tilde{z}_{ij} = \tilde{z}_{ijoverexpression} - \tilde{z}_{ij\,FLAG-GFP}$$

The resulting matrix of differences in Fisher-transformed correlation values was hierarchically clustered using the Ward method and Euclidean distances[54].

BP-R^2. Relationships between overexpression levels and markers can include non-monotonic relationships that are not properly captured with correlation metrics such as Spearman correlation. Furthermore, although the shapes induced by an overexpression were highly reproducible, the number of cells with a given expression intensity level was not. Thus, in order to quantify the strength of arbitrarily shaped relationships between markers and overexpression levels over the whole overexpression range, a density-agnostic metric termed binned pseudo R-squared (BP-R^2) was developed. For this metric, the

middle 95% of POI-GFP levels over a time-course experiment was divided into ten equal-width bins. Bins with less than 25 cells were discarded. For each bin i, the median of a measured marker ($\tilde{y}_i$) was calculated. Additionally, the overall mean of the medians of all the ten bins ($\mu_{\tilde{y}}$) was calculated. Then, for each bin, we computed the sum of squared deviations from the bin medians and the sum of squared deviations from the overall mean of medians. These values were summed over all bins, and the BP-R^2 was defined as one minus the ratio between these two values:

$$R^2_{BP} = 1 - \frac{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \tilde{y}_i)^2}{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \mu_{\tilde{y}})^2}$$

Following the rationale of classical R-squared statistics, BP-R^2 quantifies the average reduction in squared deviations per bin when modeling the data as piecewise constant within each bin (based on the bin medians) compared to using the mean over all bin medians. The BP-R^2 metric represents the relationship strength between a marker and the overexpressed signaling protein relative to the overall variability of the marker. By using the median instead of mean, the BP-R^2 selects unimodal relationships with low noise over noisy, multimodal relationships. Notably this measure works with nonlinear interaction shapes and is largely robust against density inhomogeneities. In order to aggregate the sample replicates, we considered the median BP-R^2 value across the experimental triplicates.

Threshold determination. We observed many relationships with low BP-R^2 between overexpressed proteins and measured phosphorylation markers, even within control samples (overexpression of FLAG-GFP). We propose that such weak relationships are more likely to result from indirect biological mechanisms; therefore, we focused on relationships that were stronger than all relationships seen in the controls (FLAG-GFP overexpression and untransfected cells). We chose the maximum median (across replicates) BP-R^2 of all controls (FLAG-GFP overexpression and untransfected cells) as a cutoff. Relationships that had a median BP-R^2 higher than this threshold were considered as sufficiently strong to be of interest.

Kinetic analysis. For each overexpression condition, replicates of EGF stimulation time courses were processed, stained, and measured together. Simultaneous processing enabled direct quantitative comparisons of the measured POI-GFP counts in these time courses. Samples representing all time points in a time course replicate were combined and binned by POI-GFP intensity as described in the 'Data binning' section. As the binning was performed over all samples of the same time course, the range of GFP intensity of bins with the same bin index directly corresponds to cells with similar abundance levels of POI-GFP in each of the different time points. As POI-GFP levels stay quasi-constant over the timescale of the 60-min time course (**Supplementary Fig. 8** and **Supplementary Video 1**), tracking how the median marker levels in a specific bin change over the time course reflects the kinetics of cells with a similar abundance level upon stimulation. Thus, the kinetic responses over a range of cells with low-abundance to high-abundance POI can be compared and analyzed using classical signal processing readouts such as signal response amplitude and peak time. For this analysis, only POI-GFP marker pairs with at least one strong relationship over the time course were considered.

Amplitude analysis. The response amplitude for each binned abundance level was calculated using raw counts. For each measured marker and time point, the median marker level of each POI-GFP bin was divided by the median level of the marker in the corresponding bin of the unstimulated sample (EGF 0 min) to calculate amplitude as a fold change. The amplitude for each bin was identified as the maximum fold change over all time points. Robust and strong abundance-dependent changes were identified by comparing the amplitude ratio between the second-highest and the second-lowest bin amplitude. Of those identified as robust and strong, overexpressed proteins with an amplitude ratio higher than threefold for at least two of the three replicates were identified as interesting and plotted as a heat map.

Peak-time analysis. Interesting examples of overexpression changes were defined by identifying the time point with maximum amplitude for each bin (referred to as the peak time). Consistent and robust examples were selected by the following criteria: monotonically increasing or decreasing peak times with increasing POI and a clear change in amplitude (greater than threefold) in at least two of the three replicates.

SIGNOR database comparison. To compare the consistency between the strong signaling relationships detected by BP-R^2 analysis with the relationships predicted by the SIGNOR database[34], the Python NetworkX[55] package was used to construct a sign-directed SIGNOR network from the UniProt entries of overexpressed POIs and measured phosphoproteins. NetworkX also calculates a shortest path length within the sign-directed network. Antibodies may bind to the same phosphorylation sites on more than one protein from a family, making the mapping between antibodies and UniProt entries ambiguous. In this case, the shortest path value was calculated between the overexpressed POI and any possible antibody targets.

The analysis was performed including the directionalities of signaling relationships as identified by the Spearman correlation of the bin medians in our analysis, and by exploiting the SIGNOR annotations in the following way. Simple paths between the overexpressed protein and the targeted phosphorylation sites were analyzed, starting from the shortest path to longer paths, until a sign-consistent path was found. To identify sign consistency, all edges in the SIGNOR network were classified as positive, negative, or ambiguous based on the SIGNOR 'Effect' annotation: downregulates. = negative, upregulates. = positive, something else = ambiguous. In cases for which there were multiple interaction types possible for an edge (positive and negative), the overall sign was taken to be ambiguous. In cases where the last edge was annotated to be affecting exactly the residue (SIGNOR annotation 'Residue') measured by the phospho-specific antibody through (de)phosphorylation (SIGNOR annotation 'Mechanism'), the directionality sign of this edge was determined to be phosphorylation = positive, dephosphorylation = negative or the inverse in cases the antibody was measuring the non-phospho site (e.g., Ser33/37/Thr41 on β-catenin). Measured phosphorylation sites responsible for inactivating a protein (e.g., Ser9 on GSK3β) were also signed as phosphorylation = negative. A path was determined to be sign consistent if the product of the signs of all its edges was in accordance with the relationship direction as measured by the Spearman correlation over the bins.

Systematic spillover exclusion. A stringent spillover filter was applied to systematically remove strong signaling relationships potentially affected by channel-to-channel spillover: For any measured channel that had events with ion counts over 500, we checked for spillover due to, first, isotope impurity (channels with isotopes of the same metal); second, mass resolution (–1 and +1 channels); and, third, oxidation (+16 channels). Any strong relationships (BP-R^2) with GFP and markers from these sets of channels were selected for additional verification experiments, in which the staining was done in three groups:

1. All antibodies in a set
2. All antibodies in a set except for the one that potentially causes spillover
3. Only the antibody potentially causing spillover

When spillover-induced background contributed to over 10% of the actual ion counts, the channel was discarded from further analysis (**Supplementary Fig. 12**).

Based on our spillover-exclusion protocol, we found the following channels were affected by spillover. They were excluded from the analysis performed in this manuscript, and should not be considered in any subsequent analyses using these data:

In SRC overexpression:
142Nd p-SHP2
159Tb p-SMAD1/5
In PDK1 overexpression:
142Nd p-SHP2
143Nd p-FAK
145Nd p-MAPKAPK2
160Gd p-MKK3/MKK6
162Dy p-BTK/ITK
In GSK3β overexpression:
146Nd p-p70S6K
In p90RSK overexpression:
142Nd p-SHP2
147Sm p-MKK3
162Dy p-BTK/ITK
164Dy p-SMAD2/3

doi:10.1038/nbt.3770

Data availability. All raw data are available at http://www.cytobank.org/bodenmillerlab and http://www.bodenmillerlab.org/. All data and working illustrations are available on Cytobank.org under the project 1021. The BP-R^2-based analysis is provided as **Supplementary Software** and the GitHub repository https://github.com/BodenmillerGroup/Adnet.

51. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **8**, 659–661 (2011).
52. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytometry A* **83**, 483–494 (2013).
53. Zunder, E.R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* **10**, 316–333 (2015).
54. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
55. Hagberg, A.A., Schult, D.A. & Swart, P.J. Exploring network structure, dynamics, and function using NetworkX. *Proc. 7th Python Sci. Conf.* **2008**, 11–15 (2008).